The Behavior Analyst 1985, 8, 111–119 No. 1 (Spring)

Book Review

Controlling Professional Behavior: A Review of *The Effects of Punishment on Human Behavior*[1] by Axelrod and Apsche

James M. Johnston
University of Florida

ATTITUDES TOWARD PUNISHMENT

Virtually every treatise on the topic of punishment has acknowledged widespread and fervently held cultural attitudes toward it. These attitudes vary widely from one circumstance to another, and the resulting convictions are often inconsistent and contradictory. For instance, many parents generally approve of corporal punishment in the form of paddling in the schools, but the use of electric shock under the same circumstances would easily generate outrage of headline proportions. Similarly, some people go so far as to eschew the use of punishment in the abstract while unknowingly giving and receiving the same many times every day. Others enthusiastically support the incarceration of criminals as "punishment" while preferring to "reason" with children about their "crimes."

Of course, these inconsistencies are the result of individual learning histories in particular verbal communities, but the professional verbal community of behavior analysis is rarely one of them. As a result, laypeople do not know what punishment is, how it is done, or what its effects are, and the profession of behavior analysis is forced to struggle with this ignorance in the conduct of its responsibilities.

Unfortunately, the members of the various professions that are charged with managing human behavior in different circumstances share the same cultural histories, and the extent to which the effects of these histories are modified by professional histories varies quite widely both among different professions and within the discipline of behavior analysis. It should not be surprising, therefore, that the attitudes of professionals toward punishment exhibit the same problems to some degree as do those of laypeople.

For example, many professionals hold one or more of the following statements to be generally true:

1. Punishment characteristically produces an undesirable emotional state.
2. Punishment tends to generate aggressive behavior.
3. Punishment typically produces disruption of social relations.
4. Punishment usually results in escape or avoidance behavior.
5. The punished behavior is often replaced with other inappropriate behavior.
6. Punishment tends to result in a generalized suppression of responding.
7. When a behavior is punished in one situation, it will likely increase in situations where it is not punished.

In fact, *none* of these statements is a general characteristic of the effects of properly administered punishment procedures. Nevertheless, such contentions are routinely encountered in behavior analysis and other disciplines that are faced with the need for ways to reduce or eliminate undesirable behaviors. The summary attitude toward punishment is usually a negative one, regardless of the qualifying phrases that may be attached to some misstatements, and the general conclusion often takes the form "that punishment should be used only as a last resort" (Martin & Pear, 1983, p. 207). This position is so common that one set

[1] New York: Academic Press, 1983.

of reviewers of the punishment literature complained that punishment "is actually underutilized in clinical settings" (Newsom, Favell, & Rincover, 1983, p. 285).

While the sources of these professional prejudices can ultimately be laid on the culture's doorstep, our discipline must bear considerable responsibility for failing to bring its convictions and practices under control of the available empirical facts. In fact, Skinner himself has been an occasional though consistent critic of punishment over the years. While his concerns are hardly without merit, the general attitude conveyed has unquestionably been a negative one, and many may have allowed his convictions to substitute for their own reading of the literature.

A REVIEW OF THE LITERATURE

Our responsibility to bring our professional behavior under control of the empirical literature has been made easier to fulfill by the publication of a volume edited by Axelrod and Apsche (1983) entitled, *The Effects of Punishment on Human Behavior.* Its ten chapters (338 pages) review and discuss a full range of procedures and issues concerning punishment; its author and subject indices facilitate its use, although a single reference section would have been more useful than separate and often redundant reference sections at the end of each chapter.

The book opens with an excellent, though brief, chapter by Axelrod that introduces the scope of the volume and then proceeds with a summary of the major points of each chapter. It is an easy way to see what the book is about, and the summaries tempt one to read further to discover the details.

The second chapter is by Van Houton, and it begins with a detailed discussion of issues concerning the definition of punishment. After examining problems with definitions that require two types of punishment based on presentation or withdrawal of a stimulus, he settles on "an environmental or stimulus change following instances of a specific behavior that reduce the future probability of that behavior" (p. 16) after Michael (1975). However, he proceeds further to examine indirect definitions of punishment that define it as an emotional by-product of some other process or as the presentation of a negative reinforcer. His next section is based largely on the animal laboratory literature and examines the factors influencing the effectiveness of punishment. A particularly interesting section considers the pros and cons of punishment, although a later chapter treats some of these matters in greater detail.

In Chapter 3, Van Houten is joined by Doleys to examine the literature on social reprimands (are there any other kind?). They define reprimands rather cognitively as "an expression of disapproval," an approach that is not likely to facilitate an experimental analysis of controlling variables. Given this definitional problem, it is understandable that the literature shows mixed results regarding the effectiveness of this class of "procedures," although there is more substance to this literature than one might expect.

In Chapter 4, Pazulinec, Meyerrose, and Sajwaj review the literature on response cost. They define this class of procedures as having in common the contingent removal of a positive reinforcer, and distinguish it from extinction and timeout procedures in that the latter involve the reinforcing consequence being withheld. However, this is too easy a resolution to one of our many definitional quagmires and, sure enough, a later chapter on timeout argues that the difference between response cost and timeout may only be one of the degree of stimulus change. Unfortunately, it is not just a matter of the definition of "removed" versus "withheld," but of how to encompass an enormous range of actual operations that are selected by clinical ideosyncracies unassisted by a proper analytical and technological literature. A further definitional shortcoming is that the authors do not even mention the broader sense of response cost that refers to the effort or work involved in merely emitting the response (i.e., the "cost" of

responding). Nevertheless, the chapter proceeds to review the human laboratory literature using point loss procedures, applications using tokens or money, and non-economic removals of positive reinforcers.

One of the best chapters in this volumes is by Brantner and Doherty, who consider timeout from both a conceptual and a methodological perspective. They define timeout in terms of "stimulus change in the form of change to a less reinforcing environment" (p. 88) and wrestle enthusiastically with a number of difficult issues, including distinguishing timeout from other procedures such as extinction and overcorrection. While this discussion is interesting, however, it cannot be expected to resolve matters; mere argument can only take us so far. The chapter continues with a thoughtful examination of the timeout literature in terms of the parameters influencing the procedure's effectiveness, a more useful organizational perspective than convenient categories of application.

Chapter 6 is another excellent effort—a long overdue conceptual reexamination of overcorrection, properly authored by one of its originators. Foxx and Bechtel take 87 pages to review exhaustively the massive overcorrection literature, and in the process they redefine and reinterpret this amorphous category of procedures. While they ultimately arrive at the same dilemma as do the other chapter authors (how to define one group of procedures and distinguish it from others), their clarifications offer overcorrection some badly needed respectability. For example, they take the following positions, among others: that overcorrection is a Type 1 punishment procedure, that all terms or labels (e.g., educative, functional movement training, and positive practice) should be eliminated, that it is not appropriate for aggressive individuals, that the term overcorrection has no explanatory power, and that there is no clear evidence that it differs in effectiveness from timeout and physical restraint. Of course, this brief listing does not do justice to the five recommendations and eleven clarifications in their concluding section, much less their lengthy chapter, and practitioners will greatly appreciate their thoughtful candor.

Chapter 7 by Carr and Lovaas answers three questions concerning punishment by contingent electric shock: When should shock be considered? Who will be accountable? And, who will carry it out? The chapter examines the relevant literature largely in the context of how to use shock punishment effectively and what effects to expect. It also details the evidence regarding side effects in a careful manner, in the process dispelling common misconceptions.

Chapter 8 ("Extraneous Aversives") is an oddly titled review by Bailey of studies reporting punishment applications using a miscellany of stimuli categorized under sounds, smell, tastes, visual stimuli, and "novel physical sensations." While it is only a simple survey, it is a useful compilation of this literature.

Chapter 9, by Newsom, Favell, and Rincover is another of the volume's best chapters. It carefully and dispassionately examines the literature concerning the side effects of punishment, both desirable and undesirable. After detailing the evidence regarding emotional behavior, aggressive behavior, escape and avoidance behavior, response substitution, response facilitation, generalized suppression, and punishment contrast, they conclude:

> The undesirable side effects reviewed here come from a relatively small proportion of all the studies on the therapeutic use of punishment. Even when allowance is made for the probable underreporting of negative side effects due to editorial sanctions, it is interesting to note how few studies in the literature contain observations that would suggest clinical or ethical problems. In considering the studies where undesirable side effects were observed, the overall impression that results is one of mild surprise that serious side effects are seen so infrequently. . . . Conspicuously absent from these studies is convincing evidence of serious, lasting harm to the recipients of punishment, despite the warnings of numerous authorities of various theoretical persuasions. Most of the undesirable side effects described lasted only for a few minutes or days, were quickly responsive to treatment if they did not disappear spontaneously, and constituted a relatively small and ethically justified price to pay in

return for the elimination of much more detrimental behavior. (pp. 300–301)

Their review of desirable side effects examines influences on social behavior, emotional behavior, imitation and discrimination learning, appropriate play, and attention. They conclude that

... significant benefits to the client have been reported in a number of studies, effects that are additional to the main benefit of the elimination of maladaptive behaviors. These benefits deserve fully as much consideration as undesirable side effects in decisions regarding the use of punishment. Indeed, a case can be made for according them *greater* weight than negative side effects since they tend to last longer and to make further desirable changes possible. (p. 306)

The book's final chapter by Griffith appropriately examines legal and ethical issues. The section on ethical considerations counters briefly the common arguments against using punishment, but it is less effective in this regard than some of the other chapters. Its discussion of legal matters is more useful, although it is also too brief.

SOME REMAINING PROBLEMS

In general, Axelrod and Apsche have edited a very good book. All of the chapters are well-written, and most are thorough, thoughtful, and illuminating reviews of important areas of the human punishment literature. The volume is certainly the best treatment of this literature available and its mastery should be mandatory for virtually everyone in behavior analysis and other disciplines who must manage behavior.

This book does have shortcomings, however. As is common in edited volumes, there are some overlaps and inconsistencies between chapters, although they are not especially distracting. A more serious concern is that it is quite narrowly restricted to the punishment literature and generally does not treat methods of decreasing behavior that are based primarily on other principles of conditioning. The lack of a similar treatment of these other procedures is now all the more obvious. This deficiency is particularly unfortunate because of the serious need for an integrated treatment of all methods of decreasing behavior; in fact, it is quite awkward to treat punishment-based procedures separately from others, and the authors of these chapters seemed to have occasionally chafed at this constraint.

There are a number of other substantive shortcomings of this volume, but in fairness to the editors and authors it must be acknowledged that these problems are characteristic of the punishment literature itself. Indeed, some of these problems are unfortunately characteristic of our entire technology. The complaint, then, is that the volume did not take the opportunity to address the following matters.

Definitional Issues

One of the most pervasive issues concerns the definition of procedures. The difficulty we have in defining the features and boundaries of technological procedures is at the root of a number of problems and misunderstandings, and the failure to resolve it here was quite clear. In spite of sometimes concerted efforts, virtually every chapter stumbled over this matter. Little headway was made because it is not a matter that can be resolved by reasoned argument.

Superficially, the problem is that we are uncertain about what defines our procedures so that we cannot reliably distinguish among them, and we therefore label them in an inconsistent fashion. This obviously leads to a number of related difficulties, such as improperly selected procedures, improper variations made or permitted in procedural components, and deficient training of both professionals and paraprofessionals.

The underlying causes of this problem are many, but one of them is that we have difficulty keeping principles distinct from procedures. For instance, overcorrection has been so often referred to in the context of principles that Foxx and Bechtel were forced to clarify that the term had no explanatory status and was only a procedure "composed of" behavioral principles. Of course, it has long been recognized that it does us no good that basic terms such as reinforcement and punish-

ment are routinely used to refer to both principle and procedure, but when this practice is uncritically extended in the technological arena, the risks grow rapidly.

Perhaps the confusion stems from a poor sense of what actually constitutes a technological procedure. While generically a procedure simply refers to a series of actions or operations, a formal technological procedure specifies a very particular set of actions that has been certified by experimental analysis. These actions or procedures are not explanations in any but the most superficial sense. The behavior change resulting from their application is properly explained only by the fundamental principles or organism-environment relations that describe why they work.

Now we may be getting close to the heart of the problem. In order to specify separately both a formal technological procedure, as well as its relations to principles of behavior, there must be an experimental literature that describes these relations. Here lies the fundamental weakness of our still fledgling technology. It lacks this kind of experimental base. The bulk of the punishment literature (as well as other areas of applied literature) emphasizes *demonstrating* methods of control rather than *analyzing* controlling variables. While it is indeed an experimental literature, the level of analysis is relatively superficial, usually focusing on evaluating the main effect of a complex procedure or at most on demonstrating the contribution of major procedural components. Detailed analysis of a response class, instead of a procedure, is almost rare.

As a result, punishment procedures cannot be defined by experimental evidence that clearly indicates which elements are potent, which are unnecessary, which must not be tampered with, and so forth. Instead, they are uniquely defined by each user according to his or her training history and the exigencies of each application. Then, when someone attempts to review an area of literature defined by a procedural label, the array of procedural variations thus included seems nearly limitless and inevitably extends well into what must be other procedural territories.

At some point, it becomes clear that our technology is constituted not so much of a battery of specific and distinct procedures as it is of very large, amorphous, and overlapping categories of procedural options that grow larger with every unique application. What is timeout, for example? Is it meaningful to refer to everything that goes on under that rubric as *a* procedure? Given the enormous variety of operations that are so labeled and the equivalent variety in the categories labeled "extinction" and "response cost," how can the terms not be confused?

Of course, the fundamental principles of behavior can presumably be arranged in an infinite variety of operations that are useful in managing behavior, and there is nothing wrong with lengthy lists of such procedural options as long as each is determined by sound analytical research regarding its appropriateness and effects. Research of this sort is likely to have the positive benefit of generating a sufficiently unambiguous basis for procedural definition that the resulting categories are neither amorphous nor overlapping, though they may be large depending on one's criterion. More important than niceties of nosology, such research will enable practitioners to select procedures and their options on a more informed basis with greater confidence about the outcome than is now often the case.

This kind of research, however, is only likely to emerge from a conception of technology that requires technological procedures to be based on thematic programs of analytical research. Such research must have established the mechanism of a procedure's action in the terms of the science's basic principles, specified the conditions under which it will be most effective, and described the permissible variations that will not detract from its effects. This is the conception of technology that generally characterizes other natural science technologies, and it is this kind of technological research that distinguishes real technology from tradi-

tion, experienced practice, professional folklore, or mere craft.

Comparison Studies and Generality

Most of the chapters contained either a formal section or at least an explicit discussion that attempted to compare the punishment procedure under discussion with other punishment procedures. When the dimension of comparison is administrative or logistical features of procedures, such contrasts may have some merit. However, when the comparison is in terms of the relative effectiveness of the procedures, some important misunderstandings are involved.

Comparison studies are the bane of the applied literature because they generally lead to inappropriate inferences based on evidence gathered in support of the wrong question. Their conclusions are weak for a number of reasons, only a few of which can be examined here. One of the most important considerations is that the details of the two or more procedures chosen for comparison often raise the difficult question of fairness. Because each procedure becomes the standard by which each other is measured, there must be a fair basis for comparison. Usually the similarity of their intended function is the rationale for their selection, but this obvious criterion can be misleading. For instance, even though both procedures may be properly categorized as punishment operations, any of a number of detailed differences may make a conclusion about relative effectiveness quite incorrect in many instances. If, for example, the two procedures are compared using subjects having different characteristics or in settings with different features or using parameters of administration that are less than ideal for each procedure, then the characteristics or settings or parameters of each are likely to result in effects that are misrepresentative of each category of procedures from which the present instances were chosen. The likelihood of misleading conclusions is further augmented when one of the procedures being compared is of greater interest to the experimenter than the other, which may innocently result in the favored procedure being more carefully (and effectively) conducted than the other.

One means of minimizing some of these risks requires the availability of an analytical literature that describes in adequate detail all of the variables that determine a procedure's effectiveness. Such a literature would have to make clear the details of subject and setting that dictated the selection of a particular procedure, as well as all of the parameters of its administration that were necessary for an optimum effect in any situation. Thus informed, the experimenter could then place each of the two procedures to a fair test, assured that each was conducted in its most effective manner.

However, the ability to create a fair comparison begs more fundamental questions: It is really useful to know if one procedure is more effective than another and would such an answer have much generality anyway? The importance of a shoot-out between two (or more) competitors is well-established in both lay and professional cultures. We are especially attracted to the head-to-head comparison in the social sciences because our experimental traditions are defined by the inferential statistical process of determining if there is a difference between the effects of two conditions. Nevertheless, it may be relatively rare that we need to know if one procedure is better than another.

To understand this point, it is important to appreciate that this information is only useful—indeed, it is only *possible*—when the two procedures are fully equally appropriate for the behavior change task at hand. If *anything* about the application makes one procedure more appropriate than the other, then any comparison is meaningless. (You can compare a word processor to a typewriter for the task of writing a paper, but it is meaningless to compare a word processor to a chair for the same task because they are not fully equally appropriate for that purpose.)

While it might seem easy to contend that behavior change situations are frequently addressable by more than one procedure, such arguments might be said to reflect badly on the depth and detail

of our analysis of such situations. The rejoinder might then be that while there are of course factors present in any situation that could be used to suggest one procedure over another, there is often no clear evidence that any of those factors are sufficiently critical so as to dictate unambiguously one procedure over another. However, this contention suggests the fundamental problem: We generally lack the analytical technological data-base necessary to alert us to the factors that may in fact indicate that one procedure will be more effective than another, *given that each would be optimally administered.* Without such evidence specifying the subject, setting, and administrative variables that will allow one procedure to operate more effectively than another, it is understandable that we often evaluate a particular situation as being equally appropriate for two or more different procedures. What we really need to know, then, is not the hierarchy of effectiveness of all punishment (or reinforcement) procedures, which cannot be meaningfully determined anyway, but the variables that will allow each procedure to be applied in an ideal manner, thus permitting us to recognize when one is more appropriate than any other.

There is still the matter of generality to consider. That is, is there really much generality to a comparison study, aside from these other problems? To understand why the answer to this question is in the negative, it will help to appreciate that in order for conclusions about the relative effectiveness of two or more methods to have good generality to other applications of the same methods, the critical feature of each must be known and then held constant across applications. Because the first requirement is infrequently accomplished, the second is rarely possible. Unless the central elements that define a procedure and determine its effectiveness have been identified in the existing experimental literature, they will not remain intact across its many applications, resulting in the already discussed problem of very large, amorphous, and overlapping categories of procedures. The use of simple procedural labels to refer to such categories is clearly gratuitous and implies a uniformity that is quite misleading.

Because each application under the rubric of a categorical label is likely to vary in unknown but significant ways from the particular version used in a comparison study, the relevance or meaningfulness of its conclusions for each such application is at best unknown and is more likely quite limited. Generality of method accrues not to labels but only to the particular combination of variables that make a procedure effective and that will guarantee its repeated effectiveness whenever those variables are brought together. To the extent that those variables are unknown, each application risks violating the intactness of the version used in the comparison study, thus abrogating the correctness of the prediction suggested by the comparison study. True generality of method is gained by discovering and understanding the role of critical variables whose status in particular applications can then be used to predict the effects of the version of the procedure being planned (Johnston & Pennypacker, 1980).

Finally, statements suggesting general characteristics of all punishment (or reinforcement) procedures as a whole are especially egregious for reasons that by now should be obvious. Such statements as "punishment works more quickly than reinforcement" are at the least so broad as to be wrong much of the time, and they reflect a superficial view of the technology that reminds one of its immaturity. The proper evidence to support such general statements is rarely available. The statements, then, only suggest the speaker's ignorance of the evidence that is available and our apparently inevitable desire for a technology that is sufficiently simple that it can be described in brief, declarative sentences.

SOURCES OF CONTROL OVER PROFESSIONAL BEHAVIOR

In a discussion of sources of control over scientific behavior, Moore (1981) pointed out that among the many sources of control two are especially important: "influences from operations and contacts

with data" and "influences . . . derived from the prevailing traditions of the lay community in which the scientific work is embedded" (pp. 60–61). He further reminded us that "these two systems do not exist in isolation; rather, both scientific and lay influences . . . interact conjointly to determine the behavior of the scientist . . . . The central issue . . . is how much the scientist is influenced by his contacts with operations and data relative to his contacts with the lay community" (p. 61).

Moore's analysis may be easily extended to the technological practitioner as well. However, in doing so we should appreciate that by selection, training, and professional experience practitioners may be considerably more susceptible to cultural influences than scientists. After all, the essence of science is the creation of special conditions of observation designed to maximize control by nature and minimize cultural influences. The practitioner, however, works in the cultural trenches and usually without all the protections of experimental method. This only augments the technologist's responsibility to filter his or her convictions through the finest conceptual and experimental screen that can be created by professional rigor and honesty.

The topic of punishment is one that is especially in need of this rigor. *The Effects of Punishment on Human Behavior* by Axelrod and Apsche reminds us of this need, a need that takes two particular forms. First, we must develop our technological research enterprise in general, and on special behalf of the deficiencies of the punishment literature. Our research questions must take a more thematic and analytical form so as to develop further a sound data-base that identifies the critical variables that define and distinguish punishment procedures and that enhance their proper selection and effective administration. Second, we must increase and sharpen the control that this data-base has over our technological behavior at the explicit expense of cultural sources of control. Having accomplished this, we must work to extend this control to the lay community, as we do with all facets of our science and technology.

Of course, it is difficult to be fully and accurately aware of the nature of the particular sources of control over one's convictions about punishment (or any other topic, for that matter), but there is one indication that may signal the need for an honest assessment. If it seems that those writing on this topic (such as the authors of the chapters in Axelrod and Apsche, 1983) are overly supportive of, or too passionate about, or insufficiently cautious in their evaluation of punishment procedures, then ask yourself who is more strongly and fully under control of experimental evidence regarding the appropriateness and effects of punishment procedures—you or the authors?

Some years ago I published the first review of the human punishment literature (Johnston, 1972). The present volume suggests that while the literature has grown, it has not changed. It still has the same shortcomings, but it still offers the same summary conclusions: There are available a variety of widely applicable categories of procedural operations for decreasing the frequency, duration, and other dimensions of behavior that when properly administered can be effective across a wide range of circumstances and that present a favorable balance of desirable and undesirable secondary effects. Can more be said of any other collection of procedures and their respective effects?

REFERENCES

Johnston, J. M. (1972). Punishment of human behavior. *American Psychologist, 27,* 1033–1054.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Martin, G., & Pear, J. (1983). *Behavior modification: What is it and how to do it.* Englewood Cliffs, N.J.: Prentice-Hall.

Michael, J. (1975). Positive and negative reinforcement, a distinction that is no longer necessary, or a better way to talk about bad things. In E. Ramp and G. Semb (Eds.), *Behavior analysis: Areas of research and application* (pp. 31–44). Englewood Cliffs, N.J.: Prentice-Hall.

Moore, J. (1981). On mentalism, methological behaviorism, and radical behaviorism. *Behaviorism, 9,* 55–71.

Newsom, C., Favell, J. E., & Rincover, A. (1983). Side effects of punishment. In S. Axelrod & J. Apsche (Eds.), *The effects of punishment on human behavior* (pp. 285–316). New York: Academic Press.